ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria I: PRACE MATEMATYCZNE IX (1965)

R. S. Ingarden (Wrocław)

Simplified axioms for information without probability

1. Introduction. In paper *Information without probability* by K. Urbanik and the present author (see [5] and also [4]) a system of axioms for the concept of information was given without the explicite use of the concept of probability. In the same paper the theorem was proved that if information is fully given, then probability is also uniquely determined and between the two concepts the well-known Boltzmann relation holds. Because the latter relation was used by Shannon for definition of information (see [7]), it follows from the mentioned theorem that our definition is exactly equivalent to Shannon's one as well as to all its known equivalents, among them also to axiomatic formulations using hitherto the concept of probability in an essential way (see e. g. [2], [6], [7]).

The new approach may be consider as theoretically interesting because it explains the relation between such basic concepts of science as information and probability. From the practical point of view, however, it may be argued that the new definition gives no particular advantages with respect to the previous ones, as it is much more involved and difficult to comprehend than the latter. In particular, the first two axioms proposed by us (see [5], p. 138, 139 and [4], p. 315), namely that which gives a connection between information of rings and their subrings, and that which determines a local character of information, are very hard to grasp in their intuitive meaning. The first is besides so complicated formally that even its presentation is not an easy task. After some futile, but not very penetrating trials the authors were rather skeptical about the possibility of a radical simplification of the axioms. Finally, however, in connection with a new presentation of all the theory given by the present author in a course of lectures in the University of Rochester in the fall 1962, it was found a way to a rather drastic simplification [1]. This way led not only to a simpler formulation but also, as it seems, to an easier

[1] Cf. [3]. On p. 211 there it is said that the proof of equivalence will be given in the Appendix to an article to be published in Progress in Optics, Amsterdam. Actually the article has been published in Fortschritte der Physik 12 (1964), pp. 567 - 594. Because, however, articles in Fortschritte der Physik have the review character, the author decided to publish this proof elsewhere, namely here.

approach to the intuitive understanding of the essential features of the concept of information. The aim of the present paper is to give this new presentation with all details and proofs.

To facilitate the comparison of both papers we shall follow here in principle the same formal methods as in the previous one, save some simplified notation. In order to make the present paper complete in itself we shall repeat all the needed definitions and lemmas (with their proofs) using sometimes even the same wording.

2. Notation and preliminary definitions. We shall denote Boolean rings by $X, Y, \ldots$, their elements by $x, y, \ldots, x_1, x_2, \ldots$, the zero element (considered as the same in all rings) by 0, the unit element of ring X by 1_X, the ring operations — by $+$ and $\cdot$, respectively (the latter also by a shortened version $xy = x\cdot y$). The number of atoms of ring X will be denoted by $n(X)$ or shortly n_X.

A class Ω of finite Boolean rings $X, Y, \ldots$ satisfying the conditions:

(L1) if $X \in \Omega$ and Y is a subring of X, then $Y \in \Omega$,

(L2) for any $X \in \Omega$ there exists a ring $Y \in \Omega$ such that X is a proper subring of Y,

will be called a *Boolean ladder* (or *Boolean inclusion-space*).

If on a Boolean ladder Ω a real-valued function $F(X)$ is defined, we get a decomposition of Ω into classes of equivalent rings. More precisely, we call rings X and Y *F-equivalent*, in symbols $X \underset{F}{\sim} Y$, if there exists such an isomorphism φ of X onto Y that $F(Z) = F\big(\varphi(Z)\big)$ for any subring Z of X. Moreover, we define a pseudometric on Ω by means of the following distance function between X and Y

$$\varrho_F(X, Y) = \begin{cases} 1 & \text{if } X \text{ and } Y \text{ are non-isomorphic,} \\ \dfrac{\min\limits_{\varphi}\max\limits_{Z}|F(Z)-F\big(\varphi(Z)\big)|}{1+\min\limits_{\varphi}\max\limits_{Z}|F(Z)-F\big(\varphi(Z)\big)|} & \text{if } X \text{ and } Y \text{ are isomorphic.} \end{cases}$$

We see that $\varrho_F(X, Y) = 0$ if and only if X and Y are F-equivalent.

A ring X from Ω is said to be *F-homogeneous* if for every automorphism ψ of X and every subring Y of X we have $F(Y) = F\big(\psi(Y)\big)$. The intuitive meaning of F-homogeneity is maximal uniformity of the ring with respect to function F, which means that all isomorphic subrings of X have always the same value of F.

A real-valued function $F(X)$ is called to be *regular* in a Boolean ladder provided that for any $X \in \Omega$ there exists a sequence of F-homogeneous rings $X_1, X_2, \ldots$ such that $\lim\limits_{n\to\infty}\varrho_F(X_n, X) = 0$, i.e. that the subset of F-homogeneous rings in Ω, Ω_F say, is everywhere dense in Ω.

Finally, we define two bracket notations: If $x \in X$ and $x \neq 0$, we denote by $\{x\}_X$ the subring of X consisting of all elements of X contained in x, and we call it the *inner subring* of x in X. If $x_1, x_2, \ldots, x_m \in X$, we denote by $[x_1, \ldots, x_m]_X$ the least subring of X containing all elements $x_1, \ldots, x_m$, and we call it the *outer subring* of x in X. If it is clear which a ring X we are speaking about, the subscript X may be dropped in both brackets.

3. Definition of information. A real-valued regular function $H(X)$ defined on a Boolean ladder Ω is said to be *information* on Ω if the following axioms hold:

(H1) Axiom of monotony: if Y is a proper subring of X, then

$$H(Y) < H(X). \tag{1}$$

(H2) Axiom of additivity: if X is an H-homogeneous ring and $x_1, \ldots, x_m$ are non-zero, disjoint (i.e. $x_i x_j = 0$ for $i \neq j$, $i, j = 1, \ldots, m$) its elements such that $x_1 + \ldots + x_m = 1_X$, then

$$H(X) = H([x_1, \ldots, x_m]) + \sum_{k=1}^{m} \frac{n_k}{n} H(\{x_k\}), \tag{2}$$

where $n = n(X)$ and $n_k = n(\{x_k\})$.

(H3) Axiom of indistinguishability: isomorphic H-homogeneous rings are H-equivalent.

We see that our axioms (H1) and (H3) correspond exactly to our previous axioms III and IV, respectively. Instead of axioms I and II we now have axiom (H2) which is identic with our previous Lemma 7, our previous formula (33) corresponding exactly to (2). Finally, we dropped our previous axiom V (axiom of normalization) as is frequently done in axiomatic formulations of information theory (see [6]). Indeed, in information theory the question of normalization is not so important and unique as in probability theory, and may be left open. We may use namely many different units of information (bits, nits, cal/degree, erg/degree, etc.) and it is advisable to have all formulae written covariantly with respect to change of units (dimensional homogeneity), as is customary in physics.

The intuitive meaning of our axioms is rather simple and may be explained as follows:

(i) Axiom (H1) expresses the natural property of information that it increases when the number of atoms of the respective ring increases (i.e. when the number of least details which may be yet observed grows up).

(ii) We may notice that for H-homogeneous rings the weight-factors in (2), i.e.

$$p_k = n_k/n, \tag{3}$$

are nothing else as probabilities of events x_k $(k = 1, \ldots, m)$, according to the classical Laplacean definition of probability, cf. also below, eq. (23). Therefore, we may rewrite (2) in the form

$$H(X) = H([x_1, \ldots, x_m]) + \sum_{k=1}^{n} p_k H(\{x_k\}) \tag{4}$$

which expresses the so-called *law of the broken choice* in information theory([2]). That (4) is valid not only in the case of H-homogeneous rings, i.e. when (3) is satisfied, but in the general case, can be easily proved([2]). The essential point is that it is sufficient to assume only formula (2) in which the concept of probability can be explicitely avoided.

(iii) Axiom (H3) expresses an obvious requirement that rings which are maximally uniform with respect to H cannot be distinguished from the point of view of information theory.

4. Connection between information and probability. The central role in our theory is played by the following

THEOREM. *Let H be an information on Ω. Then for every $X \in \Omega$ there exists one and only one strictly positive probability measure p_X defined on X and such that*

$$p_Y(y) = \frac{p_X(y)}{p_X(1_X)} \qquad (y \in Y) \tag{5}$$

for every subring Y of X and

$$H(X) = -\varkappa \sum_{k=1}^{n} p_X(x_k) \log p_X(x_k), \tag{6}$$

where $\varkappa$ is some positive constant and $x_1, \ldots, x_n$ are atoms of X.

It is easy to verify that for any family p_X, $X \in \Omega$, of strictly positive probability measures sytisfying (5) the function defined by (6) satisfies axioms (H1)-(H3). The H-homogeneous rings coincide with the rings having uniform probability distribution. This remark completes the interpretation of the concept of H-homogeneity.

Proof. Since information is regular on Ω, every ring from Ω is isomorphic to a subring of an H-homogeneous ring from Ω. Consequently, due to (L2) for any integer k there exists an H-homogeneous ring $X \in \Omega$ such that $n(X) \geqslant k$. Now we prove

LEMMA 1. *If X is an F-homogeneous ring from Ω and a non-zero element $x \in X$, then the subring $\{x\}_X$ is also F-homogeneous.*

([2]) Cf., e. g. [1], Chapter 2, Section 2 (Property A).

In fact, let ψ be an arbitrary automorphism in $\{x\}_X$. Putting $\psi_0(y) = \psi(xy) \cup (y \setminus x)$ we get an extension of ψ to an automorphism ψ_0 of the whole ring X. For any subring Y of $\{x\}_X$ we have $F(Y) = F(\psi_0(Y)) = F(\psi(Y))$ which implies the F-homogeneity of $\{x\}_X$.

By Lemma 1 there exists an H-homogeneous subring X_0 of X with $n(X_0) = k$. Therefore, ladder Ω contains a sequence $Z_1, Z_2, \ldots$ of H-homogeneous rings such that

$$n(Z_k) = k \quad (k = 1, 2, \ldots). \tag{7}$$

We put

$$L(k) = H(Z_k) \quad (k = 1, 2, \ldots). \tag{8}$$

Let x be an atom of Z_{k+1}. Then by Lemma 1 the ring $\{1_{Z_{k+1}} \setminus x\}_{Z_{k+1}}$ is H-homogeneous and, of course, k-atomic. Thus by (H3)

$$\{1_{Z_{k+1}} \setminus x\}_{Z_{k+1}} \underset{H}{\widetilde{\approx}} Z_k \tag{9}$$

which implies

$$H(\{1_{Z_{k+1}} \setminus x\}_{Z_{k+1}}) = L(k). \tag{10}$$

Since $\{1_{Z_{k+1}} \setminus x\}_{Z_{k+1}}$ is a proper subring of Z_{k+1}, we have by virtue of (H1)

$$L(k) < L(k+1) \quad (k = 1, 2, \ldots). \tag{11}$$

Let $x_1, \ldots, x_m$ be a system of disjoint elements of Z_{km} such that $x_1 + \ldots + x_m = 1_{Z_{km}}$ and

$$n(\{x_i\}_{Z_{km}}) = k \quad (i = 1, \ldots, m). \tag{12}$$

Now we prove

LEMMA 2. *If X is an F-homogeneous ring from Ω and $x_1, \ldots, x_m$ a system of disjoint elements of X such that $x_1 + \ldots + x_m = 1_X$ and*

$$\{x_i\}_X \underset{F}{\widetilde{\approx}} \{x_j\}_X \quad (i, j = 1, \ldots, m), \tag{13}$$

then the subring $[x_1, \ldots, x_m]_X$ is also F-homogeneous.

In fact, let ψ be an arbitrary automorphism of $[x_1, \ldots, x_m]_X$. For any index i there exist an index j_i such that $\psi(x_i) = x_{j_i}$ $(i = 1, \ldots, m)$. From (13) it follows that there exist an isomorphism φ_i of $\{x_i\}_X$ onto $\{x_{j_i}\}_X$ such that $F(Z) = F(\varphi_i(Z))$ for any subring Z of $\{x_i\}_X$. Putting for any $x \in X$

$$\psi_0(x) = \varphi_1(xx_1) \cup \varphi_2(xx_2) \cup \ldots \cup \varphi_m(xx_m) \tag{14}$$

we get an extension of ψ to an automorphism ψ_0 of X. Hence, for every subring Y of $[x_1, \ldots, x_m]_X$ we have the equality $F(Y) = F(\psi_0(Y)) = F(\psi(Y))$. Thus $[x_1, \ldots, x_m]_X$ is F-homogeneous.

We have obviously,

$$\{x_i\}_{Z_{km}} \underset{H}{\simeq} \{x_j\}_{Z_{km}} \quad (i, j = 1, \dots, m), \tag{15}$$

therefore, by Lemma 2, the ring $[x_1, \dots, x_m]_{Z_{km}}$ is H-homogeneous, and by (12) m-atomic. Thus

$$H([x_1, \dots, x_m]_{Z_{km}}) = L(m), \tag{16}$$

$$H(\{x_i\}_{Z_{km}}) = L(k) \quad (i = 1, \dots, m). \tag{17}$$

By (H2) we get

$$H(Z_{km}) = H([x_1, \dots, x_m]_{Z_{km}}) + \sum_{i=1}^{m} \frac{n(\{x_i\}_{Z_{km}})}{n(Z_{km})} H(\{x_i\}_{Z_{km}}) \tag{18}$$

or, due to (8), (16), (12), (7),

$$L(km) = L(m) + L(k) \quad (k, m = 1, 2, \dots). \tag{19}$$

It is well-known (see, e.g. [6], p. 9, 10) that every solution of equation (19) satisfying condition (11) is of the form

$$L(k) = \varkappa \log k, \tag{20}$$

where $\varkappa$ is a positive constant.

Now let X be an H-homogeneous ring and Y_0 its subring. Assume that $1_{Y_0} = 1_X$ and denote by $y_1, \dots, y_n$ all atoms of Y_0. Evidently $Y_0 = [y_1, \dots, y_n]_X$ and by Lemma 1 all rings $\{y_i\}_X$ $(i = 1, \dots, n)$ are H-homogeneous. Thus by (20) we get

$$H(X) = \varkappa \log n(X), \qquad H(\{y_i\}_X) = \varkappa \log n(\{y_i\}_X) \quad (i = 1, \dots, n). \tag{21}$$

Further, by (H2) we have

$$H(Y_0) = -\varkappa \sum_{i=1}^{n} \frac{n(\{y_i\}_X)}{n(X)} \log \frac{n(\{y_i\}_X)}{n(X)}. \tag{22}$$

Now we define a probability measure p_{Y_0} on Y_0 by

$$p_{Y_0}(x) = \frac{n(\{x\}_X)}{n(X)} \quad (x \in Y_0). \tag{23}$$

It is easy to verify that if Y_0 is a subring of Z $(Z \in \Omega_H)$, then

$$p_{Y_0}(x) = \frac{p_Z(x)}{p_Z(1_{Y_0})} \quad (x \in Y_0). \tag{24}$$

So we get from (22) and (23)

$$H(Y_0) = -\varkappa \sum_{i=1}^{n} p_{Y_0}(y_i) \log p_{Y_0}(y_i). \tag{25}$$

Thus we have defined a probability measure p_{Y_0} for any ring Y_0 from Ω_H in such a way that equations (5) and (6) hold.

Now let X be an arbitrary ring from Ω and $n(X) \geqslant 3$. By regularity of information there exists a sequence $X_1, X_2, \ldots$ of rings belonging to Ω_H such that $\lim\limits_{r\to\infty} \varrho_H(X_r, X) = 0$. We may assume that $\varrho_H(X_r, X) < 1$, i.e. that X_r $(r = 1, 2, \ldots)$ are isomorphic. By the definition of ϱ_H for every integer r there exists an isomorphism φ_r of X_r onto X such that for any subring Y of X

$$\lim_{r\to\infty} H\big(\varphi_r^{-1}(Y)\big) = H(Y). \tag{26}$$

Let $x_1, \ldots, x_n$ be all atoms of X. We shall prove that for any i $(i = 1, \ldots, n)$ the sequence

$$p_{X_r}\big(\varphi_r^{-1}(x_i)\big) \qquad (r = 1, 2, \ldots) \tag{27}$$

is convergent. Since sequence (27) is bounded, each of its subsequence- contains a convergent subsequence. Denoting the limit of such a subs sequence by $p(x_i)$, we have $\sum\limits_{i=1}^{n} p(x_i) = 1$ and $p(x_i) \geqslant 0$ $(i = 1, \ldots, n)$. Moreover, putting

$$p(x_{i_1} + x_{i_2} + \ldots + x_{i_k}) = \sum_{s=1}^{k} p(x_{i_s}) \qquad (x_{i_s} x_{i_m} = 0 \text{ for } s \neq m) \tag{28}$$

we get from (23), (25) and (26)

$$H([x, y, z]_X) = -\varkappa\,[p(x)\log p(x) + p(y)\log p(y) + p(z)\log p(z)], \tag{29}$$

where x, y, z form an arbitrary triplet of disjoint elements of X, and $x + y + z = 1_X$.

Now we prove the last

LEMMA 3. *If $X \epsilon \Omega$, $n(X) \geqslant 3$, p_1, p_2 are two strictly positive probability measures on X, and if for any triplet x, y, z of disjoint elements of X such that $x + y + z = 1_X$ we have*

$$\begin{aligned} p_1(x)\log p_1(x) + p_1(y)\log p_1(y) + p_1(z)\log p_1(z) \\ = p_2(x)\log p_2(x) + p_2(y)\log p_2(y) + p_2(z)\log p_2(z), \end{aligned} \tag{30}$$

then $p_1 = p_2$.

In fact, let us suppose, in contradiction to the statement, that there exists an atom $x_0 \epsilon X$ such that

$$p_1(x_0) \neq p_2(x_0). \tag{31}$$

Consider a triplet $x = x_0$, $y = 1_X \setminus x_0$, $z = 0$. Then by (30)

$$\begin{aligned} p_1(x_0)\log p_1(x_0) + \big(1 - p_1(x_0)\big)\log\big(1 - p_1(x_0)\big) \\ = p_2(x_0)\log p_2(x_0) + \big(1 - p_2(x_0)\big)\log\big(1 - p_2(x_0)\big). \end{aligned} \tag{32}$$

Since function $x\log x+(1-x)\log(1-x)$ $(0 \leqslant x \leqslant 1)$, is convex and symmetric with respect to $x = \frac{1}{2}$, equations (32) and (31) imply that

$$p_1(x_0) = 1-p_2(x_0). \tag{33}$$

As $n(X) \geqslant 3$, there exist such two disjoint and non-zero elements y_0 and z_0 that $y_0+z_0 = 1_X \setminus x_0$. Let us assume first that $p_1(y_0) = p_2(y_0)$ and $p_1(z_0) = p_2(z_0)$. Then we get

$$p_1(x_0) = 1-p_1(y_0)-p_1(z_0) = 1-p_2(y_0)-p_2(z_0) = p_2(x_0) \tag{34}$$

which contradicts (31). Consequently, $p_1(y_0) \neq p_2(y_0)$. A reasoning similar to that which led to (33) gives

$$p_1(y_0) = 1-p_2(y_0). \tag{35}$$

Hence and from (33) we get, according to strict positivity of p_2,

$$p_1(z_0) = 1-p_1(x_0)-p_1(y_0) = p_2(x_0)+p_2(y_0)-1 = -p_2(z_0) < 0, \tag{36}$$

which is impossible.

From (29) and Lemma 3 we imply that $p(x_i)$ is the limit of any convergent subsequence of (27). Thus, sequence (27) itself is convergent to $p(x_i)$. Setting for any element $x = x_{i_1}+x_{i_2}+\ldots+x_{i_k}$ $(x_{i_s}x_{i_m} = 0$ for $s \neq m)$

$$p_X(x) = \sum_{s=1}^{k} p(x_{i_s}), \tag{37}$$

we get a probability measure on X such that, according to (25) and (26),

$$H(X) = -\varkappa \sum_{i=1}^{n} p_X(x_i)\log p_X(x_i). \tag{38}$$

Moreover, for any subring Y of X the formula

$$p_Y(y) = \frac{p_X(y)}{p_X(1_Y)} \quad (y \in Y) \tag{39}$$

determines a probability measure on Y such that, due to (23), (25) and (26),

$$H(Y) = -\varkappa \sum_{i=1}^{m} p_Y(y_i)\log p_Y(y_i), \tag{40}$$

where $y_1, \ldots, y_m$ are all atoms of Y. Hence and from (H1) it follows that all measures p_X are strictly positive. Thus we have defined probability measures for any ring X satisfying the inequality $n(X) \geqslant 3$ and for any of its subrings. These probability measures satisfy conditions (5) and (6) and consequently by (5) every probability measure on a subring Y of X is uniquely determined by the probability measure on X. By (L2) the class of all rings X from Ω satisfying the inequality $n(X) \geqslant 3$ and all their subrings coincide with the whole class Ω. Thus to prove the uniqueness of p_X $(X \epsilon \Omega)$ it is sufficient to prove this for rings satisfying the condition $n(X) \geqslant 3$. But the last statement is a direct consequence of Lemma 3. In fact, for any triplet x, y, z of disjoint elements of such a ring satisfying the condition that $x+y+z = 1_X$ we have

$$(41) \qquad H([x, y, z)]_X) = -\varkappa\{p_X(x)\log p_X(x)+p_X(y)\log p_X(y)+p_X(z)\log p_X(z)\}.$$

On the other hand, for every probability measure $\tilde{p}_X$ on X satisfying (5) and (6) we have the same equality

$$(42) \qquad H([x, y, z]_X) = -\varkappa\{\tilde{p}_X(x)\log\tilde{p}_X(x)+p_X(y)\log\tilde{p}_X(y)+\tilde{p}_X(z)\log\tilde{p}_X(z)\}.$$

Consequently, by Lemma 3 $p_X = \tilde{p}_X$. The Theorem is thus proved.

We see that the simplification of axioms caused also an essential simplification of the proof which is more than twice shorter than before.

5. Acknowledgements. It is a pleasure to express appreciation to Professors: Alfréd Rényi of the Mathematical Institute, Hungarian Academy of Sciences, Budapest, and Joseph B. Keller of the Courant Institute of Mathematical Sciehces, New York University, New York, N. Y., for interesting discussions which made the primary impulse to the simplification of axioms of information.

For the opportunity to spend twelve months in the Courant Institute and the kind hospitality extended to him the Author thanks cordially Professor Morris Kline, head of the Division of Electromagnetic Research in the Institute, as well as the Ford Foundation and the New York University.

The author is also indebted to Professor Emil Wolf of the Department of Physics and Astronomy, University of Rochester, Rochester, N. Y. for his kind invitation to lecture as a visiting professor in the Department. As was mentioned above, these lectures gave the author an opportunity to think over the problem again.

Finally, the Author thanks cordially Professor Kazimierz Urbanik, with whom the paper [4] (used so essentially in the present investigation, especially in the proof of the central theorem) was written, and who read the present paper in manuscript and made interesting comments.

References

[1] L. Brillouin, *Science and Information Theory*, New York 1956.

[2] A. Feinstein, *Foundations of Information Theory*, New York, 1958.

[3] R. S. Ingarden, *A simplified axiomatic definition of information*, Bull. Acad. Polon. Sci., Série math., astr. et phys. 11 (1963), pp. 209-212.

[4] — and K. Urbanik, *Information as a fundamental notion of statistical physics*, Bull. Acad. Polon. Sci., Série math., astr. et phys. 9 (1961), pp. 313-316.

[5] — — *Information without probability*, Colloq. Math. 9 (1962), pp. 131-150.

[6] A. I. Khinchin, *Mathematical Foundations of Information Theory*, New York 1957.

[7] C. E. Shannon, *A Mathematical Theory of Communication*, The Bell System Technical Journal 27 (1948), pp. 379-423 and 623-656.

INSTITUTE OF PHYSICS, POLISH ACADEMY OF SCIENCES
INSTITUTE OF THEORETICAL PHYSICS, UNIVERSITY OF WROCŁAW, WROCŁAW